Non-negative Matrix Factorization for Face Recognition*

David Guillamet and Jordi Vitrià

Computer Vision Center-Dept. Informàtica
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain
{davidg, jordi}@cvc.uab.es

Abstract. The computer vision problem of face classification under several ambient and unfavorable conditions is considered in this study. Changes in expression, different lighting conditions and occlusions are the relevant factors that are studied in this present contribution. Non-negative Matrix Factorization (NMF) technique is introduced in the context of face classification and a direct comparison with Principal Component Analysis (PCA) is also analyzed. Two leading techniques in face recognition are also considered in this study noticing that NMF is able to improve these techniques when a high dimensional feature space is used. Finally, different distance metrics (L1, L2 and correlation) are evaluated in the feature space defined by NMF in order to determine the best one for this specific problem. Experiments demonstrate that the correlation is the most suitable metric for this problem.

1 Introduction

Face recognition is one of the most challenging problems to be solved in the computer vision community. Until now, several methods and sophisticated approaches have been developed in order to obtain the best recognition results using some specific face databases. Due to this huge number of methods and face databases, there is no uniform way to establish the best method because nearly all of them have been designed to work with some specific face situations. Even though, some of these methodologies have lead to the development of a great number of commercial face recognition systems. Most of the face recognition algorithms can be classified into two classes, image template based or geometry feature based. Template based methods compute a measure of correlation between new faces and a set of template models to estimate the face identity. Several well-known statistical techniques have been used to define a template model, such as Support Vector Machines (SVM) [12], Linear Discriminant Analysis (LDA) [1], Principal Component Analysis (PCA) [11] and Independent

* This work is supported by Comissionat per a Universitats i Recerca del Departament de la Presidencia de la Generalitat de Catalunya and Ministerio de Ciencia y Tecnologia grant TIC2000-0399-C02-01.

M.T. Escrig Monferrer and F. Toledo Lobo (Eds.): CCIA 2002, LNAI 2504, pp. 336–344, 2002.

Component Analysis (ICA) [2]. Usually, these approaches are focused on extracting global face features, and occlusions are difficult to handle. Geometry feature-based methods analyze explicit local facial features, and their geometric relationships. Some examples of these methods are the active shape model [4], the elastic bunch graph matching algorithm for face recognition [13] and the Local Feature Analysis (LFA) [10].

In this paper we address the problem of recognizing frontal faces captured in different illumination conditions and containing natural occlusions such as individuals wearing sunglasses and/or scarfs. In order to obtain comparable results with the most important techniques, we have used a face database that has been extensively used by the computer vision community, the AR face database [7]. Furthermore, in this paper we introduce the Non-negative Matrix Factorization (NMF) [5,6] technique in a face classification framework noticing its ability to deal with natural occlusions. As NMF is based on a subspace definition, we have also introduced the Principal Component Analysis (PCA) for a direct comparison. We also present some preliminary results concerning to the determination of which distance metric should be used in the feature space created by the positive restrictions of NMF. In order to evaluate the introduction of NMF in such a framework, we have taken as a reference the results of a previous work [3] that used the same face database for analyzing two leading commercial face recognition techniques.

2 PCA and NMF Techniques

Due to the high dimensionality of data, similarity and distance metrics are computationally expensive and some compaction of the original data is needed. Principal Component Analysis (PCA) is an optimal linear dimensionality reduction scheme with respect to the mean squared error (MSE) of the reconstruction. For a set of N training vectors $X = \{x^1, \dots x^N\}$ the mean ($\mu = \frac{1}{N}\sum_{i=1}^{N} x^i$) and covariance matrix ($\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x^i - \mu)(x^i - \mu)^T$) can be calculated. Defining a projection matrix E composed of the K eigenvectors of Σ with highest eigenvalues, the K-dimensional representation of an original, n-dimensional vector x, is given by the projection $y = E^T(x - \mu)$.

Non-negative Matrix Factorization (NMF) is a method to obtain a representation of data using non-negativity constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data [5]. Given an initial database expressed by a $n \times m$ matrix V, where each column is an n-dimensional non-negative vector of the original database (m vectors), it is possible to find two new matrices (W and H) in order to approximate the original matrix $V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu}$. The dimensions of the factorized matrices W and H are $n \times r$ and $r \times m$, respectively. Each column of matrix W contains a basis vector while each column of H contains the weights needed to approximate the corresponding column in V using the bases from W. Defining an objective function given by $F = \sum_{i=1}^{n}\sum_{\mu=1}^{m}[V_{i\mu}\log(WH)_{i\mu} - (WH)_{i\mu}]$ that is related to the likelihood of generating the images in V from the bases W and encodings H. An iterative approach to reach a local maximum of this objective function is given by the following rules [5]: $W_{ia} \leftarrow W_{ia}\sum_{\mu}\frac{V_{i\mu}}{(WH)_{i\mu}}H_{a\mu}$, $W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}}$, $H_{a\mu} \leftarrow H_{a\mu}\sum_i W_{ia}\frac{V_{i\mu}}{(WH)_{i\mu}}$. Initialization is performed using positive random initial conditions for matrices W and H. The convergence of the process is also ensured. See [5,6] for more information.

3 Experimental Results

Our experiments are based on the direct comparison of principal component analysis (PCA) and non-negative matrix factorization (NMF) techniques using a well-known face database, the AR [7]. Furthermore, the obtained results are compared to two leading techniques (see [3]) used in the computer vision community: the FaceIt and Bayesian techniques. FaceIt technique is based on the Local Feature Analysis (LFA) [10] and the Bayesian technique [9] models large non-linear variations in facial appearance using a PCA approach as a probability density estimation tool. Also, our experiments consider different distance metrics in order to find the most suitable one for the NMF technique.

The AR database [7] contains images of 116 individuals (63 males and 53 females). Original images are 768×576 pixels in size with 24-bit color resolution. This database is very interesting because subjects were recorded twice at a 2-week interval. During each session 13 conditions with varying facial expressions, illumination and occlusion were captured. Original images were reduced to 40×48 pixels because our representation becomes more manageable and faces were aligned by manually localizing both eye positions. In order to avoid external influences of background, we have only considered the part of a face inside an elliptical region. Fig. 1 shows an example of an individual taken under different conditions and the elliptical region considered.

Faces are projected in a low dimensional space that in this particular study is limited to 50, 100 and 150 dimensions in order to have a general idea of how results can change when the dimensionality of the feature space is modified. As known, Non-negative Matrix Factorization is a part based technique and Principal Component Analysis a global one and this behaviour is reflected in the bases obtained by both techniques. Fig. 2 shows some of the bases where we can initially see that NMF provides a more sparse representation instead of the global one provided by PCA.

In our experiments, training images consist of two neutral poses of each individual that were captured in two different days (instance labelled as AR 01 in Fig. 1). In order to see how each technique can deal with expressions, images labelled as AR 02, AR 03 and AR 04 are used as a testing set because they contain smile, anger and scream expressions. Table 1 shows the results of both algorithms with respect to the FaceIt and Bayesian techniques. The first impression is that L2 is not the most suitable metric when working with NMF

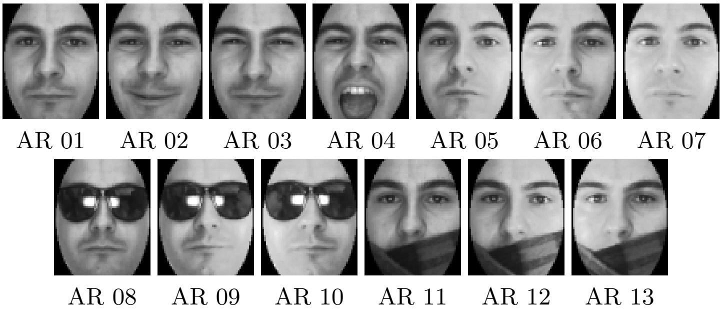
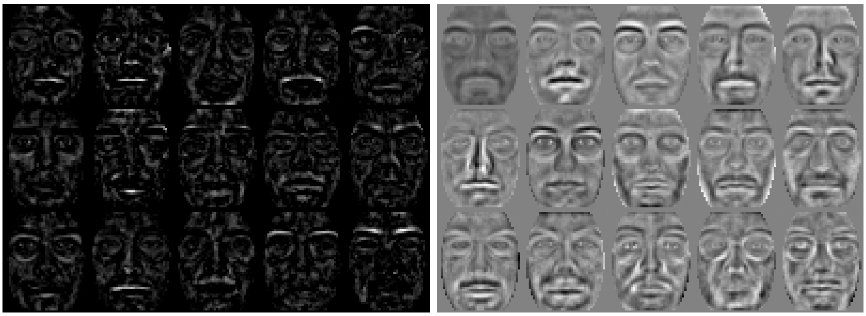


Fig. 1. Conditions of an individual of the AR face database: (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sunglasses, (9) sunglasses/left light, (10) sunglasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light



(a) NMF bases. (b) PCA bases.

Fig. 2. Bases obtained by both techniques, PCA and NMF

and both L1 and correlation metrics could be a good choice. Expression AR 02 is better classified by FaceIt and AR 03 is better classified when using NMF in a high dimensional space. But expression AR 04 demonstrates that is a very difficult expression and both PCA and NMF are not able to deal with.

Illumination conditions are also a factor to take into account in a face recognition framework. This conditions are reflected in images AR 05, AR 06 and AR 07. Table 1 shows that PCA can not deal with illumination conditions as good as the NMF and when dimensionality increases, NMF can improve the FaceIt and Bayesian approaches.

Occlusions have been considered a topic of research in the computer vision community. AR 08 contains sunglasses that occlude both eyes and AR 09 and AR 10 consider the same situation but including left and right illuminations. AR 11 images consider a scarf and that means that mouth is occluded. AR 12 and AR 13 also consider a scarf but with the addition of illumination conditions. Table 1 shows all results obtained when considering these two kinds of occlusions.

Table 1. Expression, illumination and occlusion (with sunglasses and scarf) results

	Facial Expression			Illumination Changes			Sunglasses Occlusions			Scarf Occlusions		
	AR 02	AR 03	AR 04	AR 05	AR 06	AR 07	AR 08	AR 09	AR 10	AR 11	AR 12	AR 13
FaceIt	0.96	0.93	0.78	0.95	0.93	0.86	0.10	0.08	0.06	0.81	0.73	0.71
Bayesian	0.72	0.67	0.41	0.77	0.74	0.72	0.34	0.35	0.28	0.46	0.43	0.40
PCA-50+L2 Norm	0.67	0.82	0.18	0.77	0.76	0.57	0.16	0.12	0.18	0.44	0.38	0.37
NMF-50 + L2 Norm	0.61	0.78	0.14	0.91	0.84	0.67	0.16	0.10	0.12	0.47	0.35	0.28
NMF-50 + L1 Norm	0.72	0.80	0.19	0.93	0.87	0.69	0.19	0.10	0.20	0.59	0.35	0.32
NMF-50 + Correlation	0.73	0.77	0.18	0.94	0.89	0.76	0.23	0.12	0.17	0.61	0.45	0.35
PCA-100+L2 Norm	0.80	0.88	0.24	0.86	0.86	0.69	0.23	0.15	0.22	0.59	0.50	0.47
NMF-100 + L2 Norm	0.62	0.85	0.09	0.94	0.85	0.67	0.14	0.11	0.12	0.47	0.36	0.25
NMF-100 + L1 Norm	0.85	0.91	0.29	0.97	0.94	0.87	0.24	0.15	0.21	0.66	0.55	0.46
NMF-100 + Correlation	0.89	0.90	0.28	0.99	0.94	0.88	0.32	0.19	0.24	0.76	0.62	0.59
PCA-150+L2 Norm	0.83	0.90	0.29	0.85	0.87	0.71	0.26	0.16	0.24	0.62	0.57	0.48
NMF-150 + L2 Norm	0.66	0.87	0.09	0.93	0.84	0.64	0.17	0.12	0.09	0.53	0.31	0.24
NMF-150 + L1 Norm	0.88	0.92	0.30	0.98	0.97	0.92	0.31	0.21	0.23	0.73	0.57	0.48
NMF-150 + Correlation	0.93	0.95	0.36	0.99	0.96	0.91	0.38	0.21	0.23	0.75	0.62	0.56

Under the presence of sunglasses, recognition rates decrease considerably as can be seen in table 1. This means that eyes are a very important feature to take into consideration when classifying faces. It is interesting to note that when sunglasses are considered without considering lighting influences (AR 08), NMF obtains the best recognition results. But, when lighting conditions are present, the Bayesian technique gives the best results. Thus, NMF is a good choice when partial occlusions are present but when lighting conditions affect to the scene, it turns out that NMF can not deal with a more general change in the scene. Table 1 shows a similar behaviour when considering a scarf because the NMF, when no lighting conditions are present (AR 11) in the scene, can have a high recognition rate, even comparable to the best one obtained with FaceIt, but when lighting conditions are present (AR 12 and AR 13) all recognition rates decrease considerably.

In general, the first impression of these first experiments is that NMF performs better than PCA in the same dimensional space. This behaviour was expected because PCA is based on a global transformation of the original space and NMF on a local one. Thus, it turns out that when we are considering local effects as occlusions, changes in expression or even changes in the illumination, PCA is not able to represent them as well as NMF. In terms of performances with respect to the FaceIt and Bayesian techniques, NMF has comparable recognition rates and, in some situations, is even better than these two methods. The reason of this high performance is mainly justified by its natural definition of representing data using a combination of bases that are part-based. Finally, it is clear that L2-norm is the worst metric to use with NMF and the correlation metric is the best one.

3.1 Gender Classification

It is clear that if we try to distinguish between males and females, local features corresponding to each gender are different. Thus, this means that Non-negative Matrix Factorization (NMF) can be a suitable technique for capturing these local differences. This motivates to create a gender classifier based on the NMF and when a testing face is correctly classified according to its gender, we can use this information to recognize the face using a more specific face classifier. In our study, we have learned two gender classifiers: one with PCA and the other with NMF with the same parameters as in the previous experiments. Fig. 3 shows the gender classification results when using 50,100 and 150 dimensional spaces.

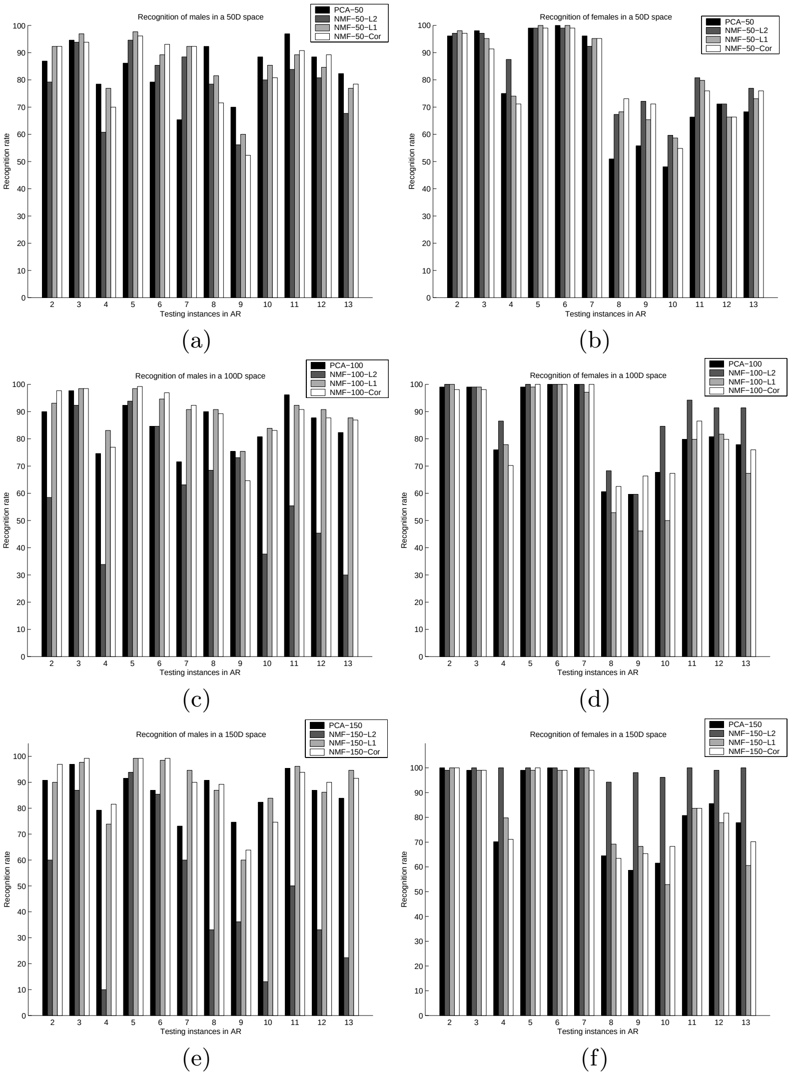


Fig. 3. Gender classification results when trying to classify males (a) and females (b) using a 50 dimensional feature space. (c) and (d) are the results when using a 100 dimensional space and (e) and (f) in a 150 dimensional space

Fig. 3 depicts a general behaviour for both PCA and NMF techniques: females are better recognized in this set of situations: AR 02, AR 03, AR 05, AR 06, AR 07 and males in the other ones. These recognition differences must be studied more deeply but this means that both genders have some local features that are better identified depending on the face situation. But images with occlusions, as AR 08, AR 09 and AR 10 are very difficult to identify even when trying to determine whether it is a male or female.

Considering that NMF is based on capturing local behaviours we can think that a more specific classifier based only on males or females should improve the initial recognition rates presented before. We have learned both PCA and NMF models for gender identification with the same internal parameters as in the previous experiments. Table 2 shows the obtained results.

Table 2. Expression, illumination and occlusion (with sunglasses and scarf) results when considering a previous gender classifier

		Facial Expression			Illumination Changes			Sunglasses Occlusions			Scarf Occlusions		
		AR 02	AR 03	AR 04	AR 05	AR 06	AR 07	AR 08	AR 09	AR 10	AR 11	AR 12	AR 13
PCA-50+L2 Norm		0.74	0.87	0.22	0.82	0.82	0.62	0.21	0.13	0.20	0.52	0.44	0.40
NMF-50 +	L2 Norm	0.68	0.83	0.17	0.91	0.86	0.68	0.20	0.10	0.14	0.51	0.34	0.31
	L1 Norm	0.81	0.87	0.25	0.94	0.89	0.76	0.24	0.14	0.22	0.60	0.41	0.37
	Correlation	0.85	0.84	0.25	0.96	0.92	0.84	0.29	0.17	0.24	0.67	0.51	0.45
PCA-100+L2 Norm		0.83	0.91	0.28	0.86	0.86	0.70	0.25	0.16	0.23	0.64	0.55	0.49
NMF-100 +	L2 Norm	0.65	0.87	0.14	0.92	0.89	0.73	0.16	0.15	0.13	0.50	0.32	0.24
	L1 Norm	0.90	0.92	0.31	0.97	0.96	0.89	0.26	0.17	0.21	0.71	0.56	0.51
	Correlation	0.91	0.94	0.34	0.98	0.97	0.92	0.35	0.21	0.25	0.79	0.62	0.57
PCA-150+L2 Norm		0.84	0.91	0.28	0.86	0.87	0.71	0.27	0.17	0.25	0.63	0.57	0.51
NMF-150 +	L2 Norm	0.70	0.88	0.13	0.94	0.89	0.69	0.18	0.13	0.10	0.53	0.36	0.27
	L1 Norm	0.90	0.93	0.33	0.98	0.98	0.92	0.32	0.20	0.24	0.73	0.58	0.52
	Correlation	0.93	0.94	0.35	0.98	0.97	0.93	0.36	0.24	0.26	0.79	0.65	0.61

In general, with the addition of a gender classifier both techniques (PCA and NMF) are slightly improved. This improvement is not very significant in face images containing hard occlusions such as those faces containing sunglasses or a scarf. However, these results motivate to build up a face classifier divided into a global gender detector and two specific face classifiers, one for males and another for females. This configuration must work out much better than only considering a face classifier because NMF is based on the representation of local features. Fig. 4 summarizes previous results showing all the recognition rates obtained according to the method used (PCA or NMF) in conjunction with their internal parameters. We have to note that the overall recognition rate for the FaceIt technique is 65.83% and 52.42% for the Bayesian one.

From the analysis of Fig. 4, we can appreciate that the introduction of a gender classifier improves the whole recognition rates even using PCA or NMF. Obviously, this behaviour is justified because it is more easy to classify a face into a male or female than recognizing the face directly. But it is clear that this improvement is more remarkable in low dimensional spaces.

If we directly compare the overall results obtained using PCA and NMF with respect to the FaceIt and Bayesian techniques, we can state that performances are comparable depending on the low dimensional space. The best configuration of our scheme is the one that uses the NMF in a 150 dimensional space using the

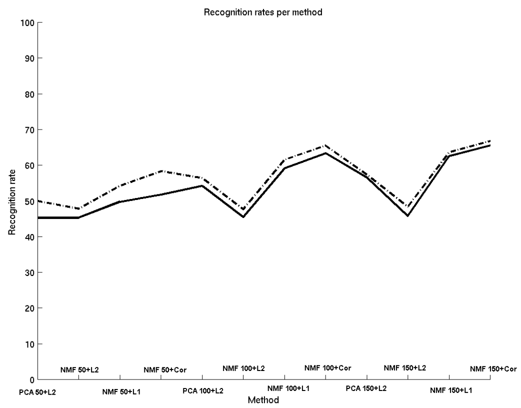


Fig. 4. Recognition rates according to the method used without a gender classifier (solid line) and with a gender classifier (dashed line)

correlation metric where we obtain a recognition rate of 66.74% that is greater than the recognition rate of 65.83% obtained by the FaceIt technique. Of course, this behaviour is not observed in all the face situations but it is clear that for some certain situations one technique will be more convenient than another one.

4 Conclusions

This paper analyzes the Non-negative Matrix Factorization (NMF) technique in the problem of recognizing faces captured under non favorable conditions such as changes in expressions, changes in the lighting conditions and occlusions. Results are also compared with the well-known Principal Component Analysis (PCA) technique because both algorithms are based on finding a subspace where our data can be expressed. Our experiments demonstrate that NMF allows high recognition rates in comparison with PCA, which treats its input data in a global a way. It is clear that these results are justified by the fact that our face database contains local variations of faces, not global ones. In this present study, we have alse analysed L1, L2 and correlation metrics noticing that the last is the most suitable one. A gender classifier stage has also been introduced in order to obtain the best results. Finally, we have compared our results to the two leading face recognition techniques (FaceIt and Bayesian), noticing that our scheme is more adapted to the problem of recognizing faces under several unfavorable conditions. This is justified because these two techniques have been designed to work with faces that contain specific changes in expressions, but not the whole range of conditions that we have exposed in this current work.

References

1. Belhumeur, P.N., Hespanha J.P., Kriegman D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):711-720, 1997
2. Comon P.: Independent component analysis - a new concept ? Signal Processing, 36:287-314, 1994
3. Gross R., Shi J., Cohn J.F.: Quo vadis Face Recognition? In: Proc. third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001
4. Lanitis A., Taylor C.J., Cootes T.: Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):743-756, 1997
5. Lee D., Seung H.: Learning the parts of objects by non-negative matrix factorization. Nature, 401:788-791, 1999
6. Lee D., Seung H.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, 2000
7. Martinez A., Benavente R.: The AR face database. Technical report 24, Computer Vision Center (CVC). Barcelona, Spain, June 1998
8. Moghaddam B.: Principal Manifolds and Probabilistic Subspaces for Visual Recognition. IEEE Trans. on Pattern Analysis and Machine Intell., 24(6), June 2002
9. Moghaddam B., Pentland A.P.: Probabilisitc visual learning for object representation. IEEE Trans. on Pattern Analysis and Machine Intell., 19(7):696-710,1997
10. Penev P., Atick J.: Local feature analysis: A general statistical theory for object representation. Neural Systems, 7:477-500, 1996
11. Turk M., Pentland A.P.: Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71-86, 1991
12. Vapnik V. N.: The nature of statistical learning theory. Springer Verlag, Heidelberg, DE, 1995
13. Wiskott L., Fellous J.M., Kruger N., Malsburg C. von der: Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(7):775-779, July 1997